教師なし深層異常検知を用いたプロトタイピング

教師なし深層異常検知を用いたプロトタイピング

Quick Prototyping Using Unsupervised Deep Anomaly Detection

江原 遥 *
Yo EHARA

Abstract : Deep Autoencoding Gaussian Mixture Model (DAGMM) [7] is a deep-learning-based anomaly detection method: unlike previous anomaly detection methods, it can be used to model high-dimensional vectors in low-dimensions. DAGMM is unsupervised: it does not require any supervision for anomalies. Instead, from the given set of high-dimensional feature vectors, DAGMM method jointly learns how to express the vectors in low dimensions and how to cluster the vectors in low dimensions. Outliers are automatically identified as vectors distant from any of the cluster centers. This feature is beneficial for quickly prototyping demonstration systems because it does not require any supervision. In this study, we show that bachelor students not familiar with deep learning can build demonstrations and analyses quickly thanks to DAGMM.

1. はじめに

教師なし深層異常検知 [7] は，近年提案された深層学習に基づく異常検知の 1 つである．[7] の従来の異常検知手法に対する利点・特徴は，次のようである．

(1) 高次元空間（例：100 次元以上）では，どのデータ点も異なるように計算されてしまう次元の呪い問題に対応するため，高次元データに対する異常検知手法である．

(2) どのデータ点が異常であるかを示す「教師データ」が不要である．

(3) 高次元ベクトルを低次元空間に射影する手法であるため，可視化に利用できる．

これらの特徴は，卒業研究のように，デモシステムをとりあえず作成してみる「プロトタイピング」の目的では有用であると考えられる．新しい研究課題に対して人工知能を適用しようとした際に生じる主な問題は，人工知能に模倣させるためのデータ（教師データ）がないという点である．したがって，卒業研究などで学生が新しい研究課題に興味を持ったとしても，人工知能に模倣させるための教師データを作成するところから始めなければならない事が多い．教師データの作成作業は，問題を理解する上で重要なステップであるという意見もあるものの，人工知能が模倣するのに十分な教師データを作成するためには，相当な時間と労力がかかり，単調作業である側面も否めない．いずれにせよ，教師データ作成のステップを短縮する事は「とりあえず，動くシステムを作ってみる」というプロトタイピングの上で，重要である．

教師データ作成の負担を軽減する方法としては，他に，教師データ作成の作業を，有償でインターネットを通じて他者に外注してしまう「クラウドソーシング」という方法もある．しかし，クラウドソーシングでは，作業者の作業能力を担保することが難しいという問題もある．また，クラウドソーシングでデータを集めたとしても，通常は最低 1 日はプロトタイピングに時間がかかる．これより早く，例えば講義のような短い時間で「とりあえずシステムを作ってみたい」という場合には，クラウドソーシングを用いたアプローチは適切とは言えない．

そこで，本稿では，試みとして，学生が興味を持った課題に対して全く教師データを用いずに，DAGMM を使ったプロトタイピングを行う手法を提案する．実際に，2 名の卒論生がこの試みに参加し，2 名とも，3 日程度でシステムを用いて分析するところまで可能になった．一方，評価用データがないため，定性的な結果しか報告できないという問題が残った．本稿では，学生の許可を得たうえで，学生が行った研究について詳述する．

関連研究

人工知能，特に機械学習分野は近年急速に発達してきたため，機械学習や深層学習を用いたシステムのプロトタイピングについては，知る限り広く普及している定番

2021 年 5 月 26 日受理
* 情報学部　コンピュータシステム学科

の手法はない．機械学習を扱うプログラミング教室などでも，実際に，自分の興味のあるテーマで深層学習を用いたシステムを作成するのに必要な期間としては，2 か月程度が準備されている [1]．実際に具体的な手法の中身まで理解するためには，数か月程度の時間が必要である事は確かであろう．近年の深層学習の手法では，有名な手法であっても，翌年には別の手法の方が良いことが判明したり，高精度に動作する理由が後から判明することも多く，手法の中身を全ての利用者に理解させることまでが必要とは必ずしも言えない．

ベースとなる手法が将来的に入れ替わることが予想されるのであれば，ベースとなる手法の定性的な傾向を迅速につかむことが望ましい．本研究では，そうしたプロトタイピングを扱う．

2. DAGMM

深層異常検知の近年の代表的な手法として，DAGMM[7] が挙げられる．DAGMM は，クラスタリング手法として有名な混合ガウスモデル（Gaussian Mixture Model, GMM）を深層化し，異常検知の機能を持たせた手法である．高次元ベクトルを次元圧縮し，低次元表現で GMM に基づくクラスタリングをした上，直感的には各クラスタ中心からの距離の和として理解できる「エネルギー値」を計算し，どのクラスタ中心からも遠い点を異常として検知する．

DAGMM は，入力ベクトル $\vec{x}$ をオートエンコーダを用いて低次元表現 $\vec{z}$ に変換し，$\vec{z}$ から $\vec{x}$ を再構成する深層学習モデルである．再構成したベクトルを $\vec{x}' = g(\vec{z}_c; \theta_d)$ とし，低次元表現を $\vec{z}_c = h(\vec{x}; \theta_e)$ とする．再構成したベクトルと元の入力の近さを測る関数を $\vec{z}_r = f(\vec{x}, \vec{x}')$ とする．ここで，この近さとしては複数の関数が利用できる．DAGMM の特徴は，低次元表現と再構成の誤差をつなげた $\vec{z} = [\vec{z}_c, \vec{z}_r]$ を最終的な潜在表現として利用することである．再構成の誤差が，潜在表現空間での距離に直接影響する．潜在表現のクラスタリングは，典型的な GMM の表記にならい，式 1で定義される．ここで，K はクラスタ数，N はデータの数，MLN は Multi Layer Network の略である．また，クラスタ k の混合係数は式 2，平均と分散共分散行列は式 3となる．

$$\mathbf{p} = MLN(\mathbf{z}; \theta_m), \hat{\gamma} = \mathrm{softmax}(\mathbf{p}) \tag{1}$$

$$\hat{\phi_k} = \sum_{i=1}^{N} \frac{\hat{\gamma}_{ik}}{N}, \tag{2}$$

$$\hat{\mu_k} = \frac{\sum_{i=1}^{N} \gamma_{ik} \vec{z}_i}{\sum_{i=1}^{N} \gamma_{ik}}, \hat{\vec{\Sigma}}_k = \frac{\sum_{i=1}^{N} \hat{\gamma_{ik}} \left(\vec{z}_i - \hat{\mu}_k\right) \left(\vec{z}_i - \hat{\mu}_k\right)^\top}{\sum_{i=1}^{N} \hat{\gamma}_{ik}} \tag{3}$$

ある入力 $\vec{x}$ の潜在表現 $\vec{z}$ について，これが異常である度合いは，式 4のエネルギー関数の値であらわされる．これは，直感的には，k 番目のクラスタの中心 $\hat{\mu}_k$ から $\hat{\vec{\Sigma}}_k$ を用いて $\vec{z}$ への距離を測り，全クラスタからの距離の和が大きい $\vec{z}$ を異常と判定していると解釈できる．もちろん,「異常」の比率はデータに依存する．[7] の例では，単純に，エネルギー値上位 20% を異常と判定している．

$$E(\vec{z}) = -\log \left(\sum_{k=1}^{K} \hat{\phi}_k \frac{\exp\left(-\frac{1}{2} \left(\vec{z} - \hat{\mu}_k\right)^\top \hat{\vec{\Sigma}}_k^{-1} \left(\vec{z} - \hat{\mu}_k\right)\right)}{\sqrt{\left|2\pi \hat{\vec{\Sigma}}_k\right|}} \right) \tag{4}$$

訓練は，ニューラルネットワークのパラメタ $\theta_e, \theta_d, \theta_m$ に対して，下記の目的関数を最小化することで行う．L はベクトルの再構成に関する損失関数，P は罰則項であり，λ はハイパーパラメタである．

$$J(\theta_e, \theta_d, \theta_m) = \frac{1}{N} \sum_{i=1}^{N} L(\vec{x}_i, \vec{x}_i') + \frac{\lambda_1}{N} \sum_{i=1}^{N} E(\vec{z}_i) + \lambda_2 P(\hat{\Sigma}) \tag{5}$$

3. プロトタイピングのテーマ

DAGMM は高次元の異常検知手法であるが，まずは，高次元ベクトルに対して異常検知を行いたい動機を卒研生に理解させる必要があった．そこで，8 名の卒研生全員に対して，Bidirectional Encoder Representations from Transformers (BERT) [6] を用いた，テキストから，テキストの意味を表現する空間への変換の例を 30 分程度説明した．その後，深層異常検知の基本的な動作を 30 分程度で概説し，テキストの意味的な外れ値の同定が BERT と DAGMM を組み合わせることで実現できるかもしれないこと，DAGMM の自然言語処理への応用についてはほとんど先行研究がない事を説明した．その後の相談の結果，2 名の卒研生が本研究の深層異常検知を用いた研究を行いたいと申し出た．

1 名は,「小説家になろう」のテキストから，意味的な外れ値となる文を探すことで，特徴的な表現が抽出できないかというアイデアを提案した．1 名は，英語読解のためのテキスト集合から，意味的な外れ値となる文を探す研究テーマを申し出た．

4. 「小説家になろう」のプロトタイピング実験内容

タイトルデータデータは「なろう分析記録 [2]」の「『なろう小説 API』を用いて、なろうの『全作品情報データを一括取得する』Python スクリプト※コード改良しました」という記事にて紹介されているプログラムを利用した．「小説家になろう」累計人気順の上位 10 作品を取得した（表 1）．なお文章については追記や作者の一言など本篇に関係のない部分や 20 文字以上の長い行を取り除いた．

各小説に対して，$k = 2$ の DAGMM を適用し，2 次元で可視化した結果が図 1である．$k = 2$ とした理由は，異

表 1: 本文データ行数．タイトルは，https://syosetu.com/より引用．

タイトル	空白を除く全行数	除去後の行数
転生したらスライムだった件	70601	20910
とんでもスキルで異世界放浪メシ	64216	23819
無職転生　- 異世界行ったら本気だす -	134136	68896
ありふれた職業で世界最強	87435	22331
Ｒｅ：ゼロから始める異世界生活	131594	25045
デスマーチからはじまる異世界狂想曲	104804	32188
ヘルモード　～やり込み好きのゲーマーは廃設定の異世界で無双する～	41458	12415
陰の実力者になりたくて！	17923	8436
八男って、それはないでしょう！	84932	23240
蜘蛛ですが、なにか？	56015	26578

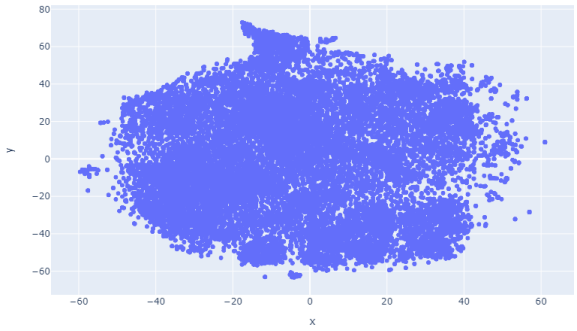

図 1: 転生したらスライムだった件

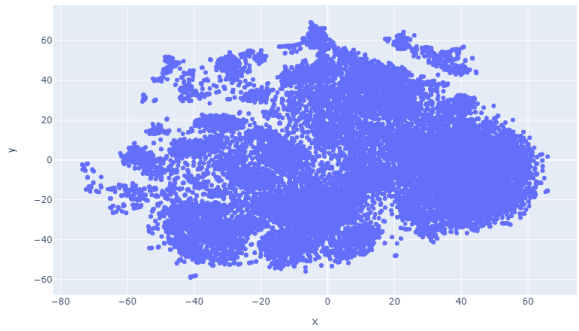

図 2: とんでもスキルで異世界放浪メシ

表 2: 「転生したらスライムだった件」異常値上位
文章はhttps://ncode.syosetu.com/n6316bn/より引用．

順位	値	文章
1	0.81996584	子供だな。
2	0.81976426	暴風、破壊、腐食、滅び。
3	0.81943774	「勿体無きお言葉――」
4	0.8194159	うん。言わなくても解ってる。
5	0.819387	変化は唐突であり、苛烈だった。
6	0.8193684	それに、理由はそれだけではなく……。
7	0.81933355	その攻撃により、近藤は致命傷を負う。
8	0.81930774	しかしそれも一時凌ぎであった。
9	0.8192735	第 12 試合…… ハクロウ ｖｓ シオン
10	0.8192479	お前、50 万個って、ひょっとして……

常値と正常値の 2 クラスタに分類する方が解釈が容易であると考えたためである．また，少なくとも単語の意味的多様性については $k = 2$ で十分であるという報告があることも理由の 1 つである [5]．

この図 1は主に大きい二つのクラスタが確認できる．下部に存在する小さいクラスタは会話，上部に存在する大きなクラスタはそれ以外といった様子である．セリフのクラスタで多いものは疑問符・感嘆符が最後にあるクラスタが多いようである．さらに小さいが登場人物の叫びと擬音語のクラスタなどもある．そしてロールプレイングゲームに用いられる単語のクラスタも存在している．

表 2の上位の結果を見るとセリフの割合はあまり多くなく，登場人物の心の中の言葉，そして場面の説明の中で，並列など特徴的な表現をしているものが抽出できているように見える．ロールプレイングゲームに用いられる単語の異常値はそこまで高くないことが分かる．

このように，各小説に対して，DAGMM を用いた可視化を行い，定性的な性質を調べた．その結果を下記に述べる．図のキャプションは，作品名である．

図 2はあまりクラスタがまとまっていないものが多い．つまり様々な特徴を持った文章がある作品だということがわかる．このクラスタにもロールプレイングゲームに用いられる単語のクラスタ，登場人物の叫びや擬音語のクラスタは存在している．しかし「転生したらスライムだった件」のクラスタの大きさよりも大きいクラスタであった．

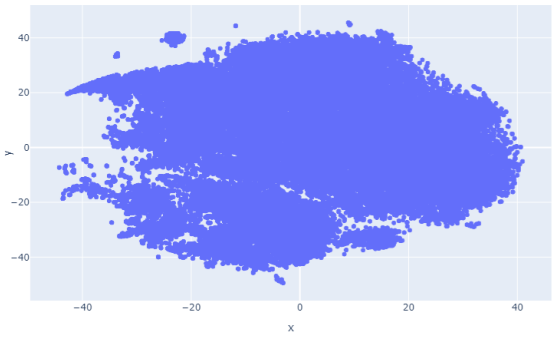

図 3: 無職転生　- 異世界行ったら本気だす -

この図3は上下に分かれている大きなクラスタと，小さなクラスタが見て取れる．上部のクラスタはセリフ以

外を多く含むクラスタ，下部のクラスタはセリフ多く含むクラスタである．

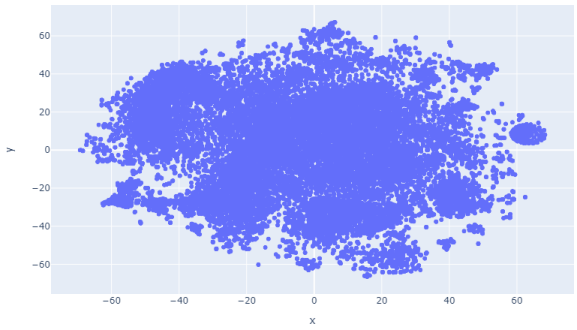

図 4: ありふれた職業で世界最強-

図 4の約 7 割がセリフのクラスタであり, 右端には楕円型の無言のクラスタが存在している．

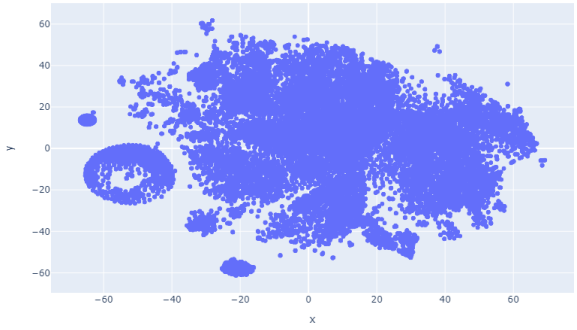

図 5: Ｒｅ：ゼロから始める異世界生活

図5のクラスタは大きな楕円型のクラスタが特徴である．

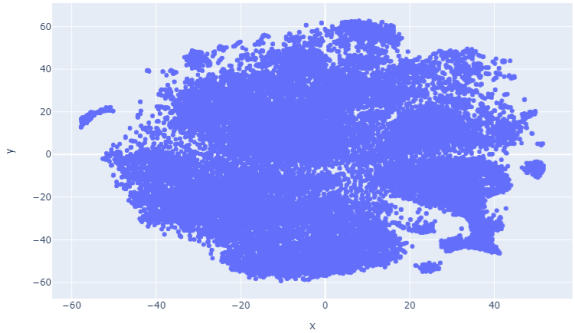

図 6: デスマーチからはじまる異世界狂想曲

図6の大きなクラスタは約 $y = 0$ 軸を境界にして上部がセリフ，下部がセリフ以外で分かれている．左側の外れた位置には英単語を多く含んだクラスタが存在している．

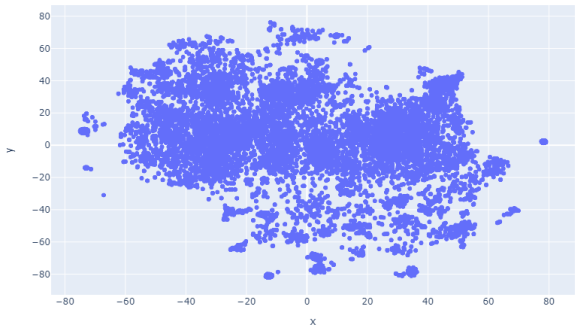

図 7: ヘルモード　～やり込み好きのゲーマーは廃設定の異世界で無双する～

図7の中央のクラスタは約 $x = 0$ 軸を境界として左側がセリフの集合，右側がそれ以外の集合である. その大きなクラスタの下部には多くの小さなクラスタが存在している．

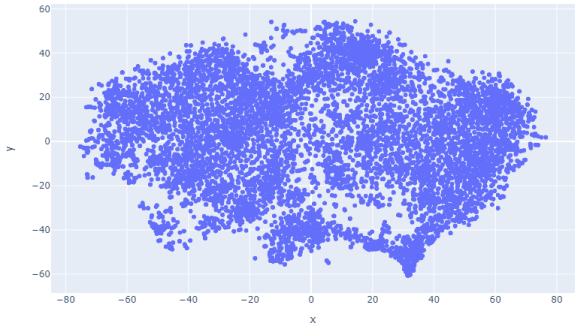

図 8: 陰の実力者になりたくて！

図8のクラスタは約 $y = 0$ 軸を境界として右側がセリフ，左側がそれ以外のクラスタとなる．

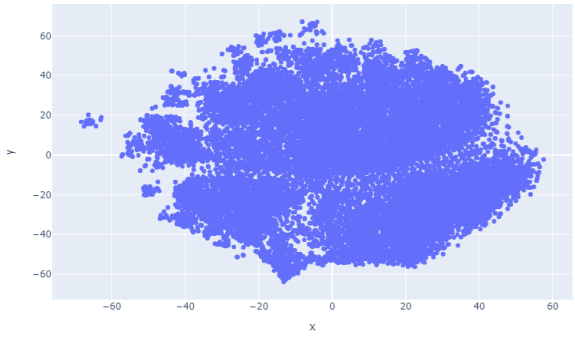

図 9: 八男って、それはないでしょう！

図9では約 $y = -8$ から約 $y = 58$, 約 $x = -56$ から約 $x = 5$ の間に存在しているクラスタがセリフ以外である．それ以外のクラスタはセリフである．約 7 割がセリフであることがわかる．

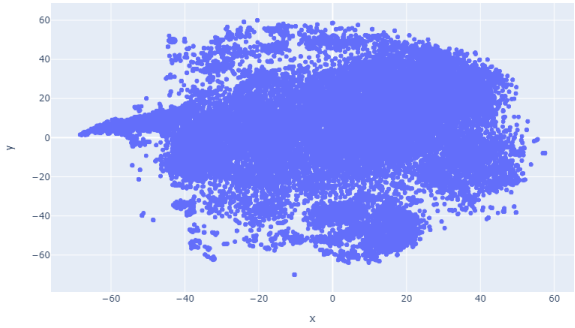

図 10: 蜘蛛ですが、なにか？

図10では，最も大きなクラスタはセリフ以外のクラスタである．また，約 2 割がセリフのクラスタである．この作品にはロールプレイングゲームに用いられる単語のクラスタも存在している．さらに登場人物の心の声のクラスタがあるという特長がみてとれる．

4.1 会話率について

会話率とは「なろう API」にて計算されている会話の割合である．求め方は「会話率 - なろうデベロッパー [3]」によれば，次のとおりである：“会話行÷ (地の文+会話行) × 100 で会話比率を求めています。” 式の中にある会話行とは “「” で始まり “」” で終わる一つの行のこと定義されている．

表 3: 異常度の高い文例．Extensive Reading Central より文例を引用．

異常度	文例
9.156665	Well, if you are one of these people, try to get through this thought.
4.7371583	You see, this is a problem that you will go through one way or the other.
2.2981603	It can happen to teachers.

タイトル	会話率
転生したらスライムだった件	14
とんでもスキルで異世界放浪メシ	36
無職転生　- 異世界行ったら本気だす -	22
ありふれた職業で世界最強	41
Ｒｅ：ゼロから始める異世界生活	40
デスマーチからはじまる異世界狂想曲	38
ヘルモード　～やり込み好きのゲーマーは廃設定の異世界で無双する～	28
陰の実力者になりたくて！	39
八男って、それはないでしょう！	41

図 11: 会話率．タイトルは [4] より引用．

解析した一部の作品のセリフのクラスタの割合がかなり高かったが，元々の会話率の図 11を見てみると多くても半分以下の割合であることがわかる．セリフの割合が高くなってしまった理由については，文字数が 20 文字以上になってしまい除外された文章にあまり会話文が含まれていなかったことが考えられる．

5. 英語教材に対するプロトタイピングの実験内容

まず，文献 [8] の再現実験を行ってもらった．次に，静岡理工科大学で用いている英語多読教材（Extensive Reading Central, `https://www.er-central.com/`）のうち Level 5 の 1 文書に対して BERT と DAGMM を適用し，異常度の高い文例を表 3に，列挙した．この表の各文は，単語頻度を用いた難易度推定では，同程度の難度の例文であるが，実際の異常度の値には大きな違いがある．おそらく，テキスト中で他の文にない表現が抽出されていることが示唆される．

6. おわりに

本稿では，卒研生 2 名に対して，深層異常検知を用いたプロトタイピングを行ってもらった．どちらの学生も，数日程度で，自分から深層異常検知を用いた分析を行えるようになった．一方，教師データを作成する必要性がない点はスムーズに進んだ一方，定性的な結果にとどまり，結果の解釈が難しい問題が残った．小説の分析においては，セリフとセリフ以外を峻別する程度の可視化はできたものの，それ以上の結果については，解釈が難しい．将来的には，教師データを使わないことにこだわるのではなく，研究テーマに対して必要な教師データ作成

の作業量を見積もることが可能となることが望ましいと思われる．

謝辞

本研究は静岡理工科大学 2020 年度提案型教育研究費研究プロジェクト,「「やらまいか！」を醸成する人工知能応用のための教師なし深層異常検知技術基盤の開発」の支援を受けたものである．また，本研究は，科学技術振興機構 ACT-X 研究費 (JPMJAX2006)，ならびに日本学術振興会科学技術研究費補助金（18K18118）の支援も受けた．卒業研究のデータ等の提供を申し出ていただいた望月一輝氏，古奈広隆氏に感謝する．また，英語多読教材のデータを提供していただいた静岡理工科大学の谷口ジョイ先生，並びにそのデータ入力・整形などの作業に協力していただいた谷口ジョイ先生の学生の方に感謝する．

参考文献

1) https://ledge.ai/2019-04-08-333495545caacf2446015/.
2) https://karupoimou.hatenablog.com/.
3) なろうデベロッパー. https://dev.syosetu.com/man/kaiwa/.
4) 小説家になろう. https://syosetu.com/.
5) Ben Athiwaratkun, Andrew Wilson, and Anima Anandkumar. Probabilistic FastText for multi-sense word embeddings. In *Proc. of ACL*, pp. 1–11, Melbourne, Australia, July 2018. Association for Computational Linguistics.
6) Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *Proc. of NAACL*, pp. 4171–4186, Minneapolis, Minnesota, June 2019.
7) Bo Zong, Qi Song, Martin Renqiang Min, Wei Cheng, Cristian Lumezanu, Daeki Cho, and Haifeng Chen. Deep autoencoding gaussian mixture model for unsupervised anomaly detection. In *In Proc. of ICLR*, 2018.
8) 江原遥. 外国語語彙学習のための教師なし深層異常検知に基づく語の用例の多義性・主要性の提示. 人工知能学会全国大会論文集, Vol. JSAI2020, pp. 2D1GS902–2D1GS902, 2020.